中華農業研究（Jour. Agric. Res. China）47(1)：12～27（1998）

分段直線迴歸模式與傳統迴歸法在評估作物對環境之非線性反應的比較研究[1]

呂　　秀　　英[2]

摘要：一個高適應性的作物品種，應該要對惡劣環境鈍感而對有利環境敏感，亦即這種對環境優劣表現互異的理想品種，事實上對環境呈非線性反應。傳統上常用直線迴歸分析來評價基因型的穩定性，是假設基因型對所有環境具直線反應，當參試品種族群中存在有對環境優劣表現差異的基因型時，將會產生遮蓋效應而無法選出眞正理想的品種，甚至將不穩定的品種誤認爲是穩定。因此，本研究擬探討藉 Verma and Chahal (1978) 所提出之分段直線迴歸來進行穩定性分析的可行性。以玉米及水稻假設模式爲基礎，各產生200套的作物區域試驗模擬資料後，再將每一套數據用傳統迴歸法及分段直線迴歸法分別進行穩定性分析，以比較這兩種方法的估計效率。結果發現，利用分段直線迴歸法較傳統迴歸法可有效地檢測出品系對優劣環境之個別反應差異，同時用來選拔出高適應性的作物品種，可減少誤判的風險。本研究並以個人電腦之 SAS 軟體，編寫出一個穩定性分段直線迴歸分析程式，以本省玉米區域試驗之實例資料來說明其進行分析的流程及結果，幫助農業工作者之實際應用。最後，利用一個三度空間圖來描述各品系在優劣環境之穩定性與平均表現值間的關係，將有利於篩選出具高產且高適應性的基因型。

關鍵詞：穩定性、分段直線迴歸、非線性反應、模擬試驗、假設模式、SAS 程式。

前　　言

有關穩定性大小的評估方法，迄今各方學者所提出者不下數十種(1,5,6,7,13)，其中以直線迴歸分析法(3,4,12)最爲常見。該法首先是以各環境之平均產量與整個試驗的總平均產量之偏差，作爲該一試驗的環境效應估值，稱爲環境指標（environmental index）(4)，然後將各品種（基因型）在各環境之產量隨這些指標作迴歸分析，並以最小平方法（least square method）估計迴歸係數，利用這些迴歸係數估值（b）可以評價品種的穩定性。在本省重要作物育種實施程序中，一直是利用直線迴歸法來分析區域試驗的品種穩定性，以選出相對於推廣品種之高產且穩定的品系。直線迴歸穩定性分析法具有計算簡單而容易解釋的優點，然其使用效益迄今仍備受爭議(1,5,7,13,14,15,16)。各方學者大多針對數理上的前提限制及應用範圍來加以討論，而本研究只想就實際層面對目前常用的直線迴歸穩定性分析法來提出質疑：該法眞的能選出眞正理想的品種嗎？因爲一個眞正符合農業觀念的理想品種，應該是在惡劣環境下仍能保持高產且穩定的表現，但在有利環境下可發揮其最大的產量增進能力，換言之，這種對惡劣環境鈍感而對有利環境敏感的品種，具有最高的適應性，爲農業育種選拔上最理想的基

---

1. 臺灣省農業試驗所研究報告第 1927 號。
2. 本所農藝系副研究員。臺灣省　臺中縣　霧峰鄉。

因型，其對環境優劣的表現互異，即對所有環境的反應事實上呈一種非線性關係。然而，傳統的直線迴歸穩定性分析法，是在假設所有基因型對環境均具直線反應之前提下，將各基因型對所有環境指標作迴歸分析，當品種在惡劣環境及有利環境的表現確實存有反向差異時，很有可能會產生互相抵銷的遮蓋效應（masking effect），以致無法選出眞正理想的品種。對此，Verma and Chahal[14]首先將環境指標分割爲正負兩個子集合，然後再將各基因型分別對該兩個環境子集合進行傳統直線迴歸分析，由兩條迴歸線所估得的兩個迴歸係數各表示基因型在優劣環境下的穩定性係數。該法在觀念上相當簡單且實用，同時由於品種對優劣環境之穩定性評估結果的等級相關（rank correlation）並不顯著，因此將所有環境指標分割成兩部分，再分別進行迴歸分析，其彼此間的評估結果將不會互相影響[14]。但是直線迴歸法應用於穩定性分析上，有其根本之限制，即基因型及環境數不得太少[5,6]，因此當有限的環境數被再分割成兩個子集合時，將勢必造成環境數會變得更少，而影響各子集合分別所估得之迴歸線斜率的顯著性，以致降低了穩定性分析的準確度[14]。換言之，每一基因型都分別估算兩條迴歸線，需要更多的環境或重複數[13]。事實上 Verma and Chahal 的方法，與假設兩段迴歸直線間之結點（knot）已知爲 0 的分段直線迴歸模式（piecewise linear regression model）[11]之觀念完全相同，只不過後者是以含有兩個迴歸係數的單一個複迴歸方程式的形式來表達而已；經由分段直線迴歸模式所求配的參數估計值，理論上與 Verma and Chahal 法所得者相同，但多少可避免因爲分開估算兩條迴歸線之情形下所導致之直線不易顯著的問題。

本研究將利用模擬試驗，比較分段直線迴歸模式與傳統直線迴歸模式所估得的穩定性係數之估值取樣分布的平均、變方及其估計偏差，以評估分段直線迴歸模式在穩定性分析的可行性。繼之，爲幫助農業工作者之實際應用，本研究並提供一個以 SAS 軟體編寫的穩定性分段直線迴歸分析程式，利用本省玉米區域試驗的實際資料爲例，說明其進行分析的流程及結果。

材料及方法

一、穩定性分析的方法

㈠傳統直線迴歸法[4]

基因型（品種）i（i=1…k）在環境 j（j=1…n）下的平均性狀觀測值 Y_{ij} 以直線迴歸模式表示：

$$Y_{ij}=\mu_i+\beta_i E_j+\varepsilon_{ij}$$

其中 E_j 爲環境 j 的效應，即環境指標，其估計式爲

$$\hat{E}_j=\bar{Y}_{.j}-\bar{Y}_{..}$$

而

$$\bar{Y}_{.j}=\sum_{i=1}^{k}Y_{ij}/k, \quad \bar{Y}_{..}=\sum_{i=1}^{k}\sum_{j=1}^{n}Y_{ij}/kn$$

；Y_{ij} 隨環境指標作迴歸，利用最小平方法可估得截距 μ_i 之估計值 m_i，以及迴歸係數 β_i 之估計值 b_i；而 ε_{ij} 爲迴歸離差項，滿足平均爲 0 及變方爲 σ^2 之常態分布。

利用迴歸係數估值來評價品種的穩定性：當 b=1 時，表示品種爲符合農業觀念的穩定性，其會隨著環境改善而表現增進，但也不致因環境不利而表現太差；當 b>1 時，爲不穩定的品種；當 b=0 時，則表示該品種不論環境如何變化，均能維持其一致的表現，屬生物觀念的穩定性，此多應用於病蟲害防治的品種選拔。

傳統迴歸穩定性分析法之計算，可參考呂[10]所編寫的 SAS 程式。

㈡分段直線迴歸法[14]

假設 E_x 爲連接兩段迴歸直線的已知結點，則基因型 i 隨環境指標 E_j 的分段直線迴歸

分析模式爲：

$$Y_{ij}=\mu_i+\beta_{1i}E_j+\beta_{2i}(E_j-E_x)I_j+\varepsilon_{ij}$$

其中

$$I_j=\begin{cases}0 & \text{if } E_j\leq E_x\\ 1 & \text{otherwise}\end{cases}$$

設定 E_x 爲0，即將環境指標劃分爲以下兩部分：

1. 惡劣環境指標

$E_j\leq 0$，其 $I_j=0$，即迴歸模式爲

$$Y_{ij}=\mu_i+\beta_{1i}E_j+\varepsilon_{ij}$$

2. 有利環境指標

$E_j\geq 0$，其 $I_j=1$，即迴歸模式爲

$$Y_{ij}=\mu_i+(\beta_{1i}+\beta_{2i})E_j+\varepsilon_{ij}$$

其中 β_{1i} 及 $(\beta_{1i}+\beta_{2i})$ 各爲品系 i 之性狀觀測值分別隨惡劣環境指標和有利環境指標之迴歸係數，因此各以 β_p 及 β_f 之符號代之，其估計值 b_p 及 b_f 用來表示基因型分別在惡劣及有利環境下的穩定性係數。

根據 b_p 及 b_f，基因型在不同環境下的相異表現有九種可能[14]，如表 1 所示：其中第 I 類爲無論環境如何變化均表現最穩定的基因型，這種穩定性屬於生物觀念的穩定性，對農業上育成高產品種之改良幫助不大，但有利於病蟲害品種選拔；第 II 、III 類基因型在惡劣環境下表現穩定，而在有利環境下則產量可獲增進，其增進效果完全決定於基因型平均值之表現，當平均值同時具有高產表現時爲農業品種選拔上最理想的基因型，因爲當環境改善的程度愈好，其表現也愈好，但在逆境時則仍能維持一定的水準，但當第 II 、III 類基因型不具有高產表現時，表示其在有利環境下的增進效果不如其他基因型；反之，第 VII 類基因型在惡劣環境下表現敏感，但在環境變好時卻未見產量有明顯改善，此爲最不好的基因型；第 IX 類爲無論環境如何變化均最爲敏感，是最不穩定的基因型；至於其他種類的基因型，其表現則介於上述各種類別之間。

表 1. 根據分段直線迴歸法之優劣境穩定性係數估值進行基因型適應性之分類

Table 1. Classification of adaptation of genotypes on the basis of stability indices from piecewise linear regression in poor and favorable environments

Categoty	Stability index[z]		Adaptation	
	b_p	b_f	Poor environments	Favorable environments
I	<1.0	<1.0	high stability	high stability
II		=1.0	high stability	average stability
III		>1.0	high stability	high sensitivity
IV	=1.0	<1.0	average stability	high stability
V		=1.0	average stability	average stability
VI		>1.0	average stability	high sensitivity
VII	>1.0	<1.0	high sensitivity	high stability
VIII		=1.0	high sensitivity	average stability
IX		>1.0	high sensitivity	high sensitivity

[z] b_p,b_f : Stability indices in poor environments and favorable environments respectively, from piecewise linear regression.

分段直線迴歸穩定性分析法之計算，以 SAS 套裝軟體中最基本的 STAT 功能系統[17]編寫程式如附錄 1。

二、模擬試驗的方法

由於很難找到一個實際區域試驗資料能包含表 1 所述的各種基因型，爲組配成如表 1 的所有可能之基因型種類，因此本研究以給定迴歸係數眞值的假設模式（hypothetical model）爲基礎進行模擬試驗，才能客觀地探討各種求配程序所估得的迴歸係數估值之取樣分布的特性。本研究選定兩種不同作物之區域試驗的假設模式：

㈠玉米假設模式

假設有12個基因型爲參試品系，在13種相異環境下各呈現不同的適應性反應類型，由 Lu and Tien[9] 之飼料玉米綜合品系秋作區域試驗的觀測值分布特性（平均值爲 2942 kg ha^{-1}，標準偏差爲474），所假設的各品系在優劣環境下之穩定性係數、各品系之平均產量及環境指標眞值列如表 2 。

表 2. 模擬試驗之假設模式所採取的各品系對優劣環境之線性反應、平均產量及環境指標的眞值

Table 2. True values of linear responses of genotypes to the poor and favorable environments, mean yield and environmental indices used as hypothetical models in the simulation experiments

Maize model (σ=474)[w]					Rice model (σ=392)				
Real linear response of genotype[x] (12)[y]			Mean yield (kg ha^{-1})	Envir index (13)	Real linear response of genotype[x] (15)			Mean yield (kg ha^{-1})	Envir index (10)
Category[z]	β_p	β_f			Category	β_p	β_f		
I	0.2	0.5	3100	−1200	I	0.3	0.3	6100	−1300
II	0.2	1.0	2700	−1000	II	0.3	1.0	6400	−1000
III	0.2	1.5	2900	−800	III	0.3	1.7	6700	−700
I	0.6	0.5	2600	−600	I	0.6	0.3	6000	−400
II	0.6	1.0	2800	−400	II	0.6	1.0	6800	−100
III	0.6	1.5	3100	−200	III	0.6	1.7	6300	200
IV	1.0	0.5	2800	0	IV	0.9	0.3	6000	500
V	1.0	1.0	3200	200	V	0.9	1.0	6300	800
VI	1.0	1.5	3500	400	VI	0.9	1.7	6500	1100
VII	1.4	0.5	2700	600	VII	1.2	0.3	5700	1400
VIII	1.4	1.0	2500	800	VIII	1.2	1.0	6300	
IX	1.4	1.5	3300	1000	IX	1.2	1.7	5900	
				1200	VII	1.5	0.3	5400	
					VIII	1.5	1.0	6200	
					IX	1.5	1.7	6600	

w σ : Square root of variance of the hypothetical model.
x β_p, β_f : True values of regression coefficients in poor environments and favorable environments, respectively.
y Number of genotypes or environments given in parenthesis.
z See Table 1.

㈡粳稻品系的假設模式

假設有15個基因型爲參試品系，在10種相異環境下各呈現不同的適應性反應類型，由 Lu *et al.*[8] 之粳稻品系第一期作區域試驗的觀測值分布特性(平均值爲 6298 kg ha^{-1}，標準偏差爲392)，各品系在優劣環境下之穩定性係數、各品系之平均產量及環境指標眞值列如表 2。

繼之，藉 SAS 統計軟體所提供之標準常態逢機變數生產函數 RANNOR[17] 來產生區域試驗的逢機變值：

$$Y_{ij}=\begin{cases}\mu_i+\beta_{pi}E_j+\sigma\times \mathrm{RANNOR(seed)} & \textit{if } E_j\leq 0\\ \mu_i+\beta_{fi}E_j+\sigma\times \mathrm{RANNOR(seed)} & \textit{if } E_j\geq 0\end{cases}$$

式中σ爲試驗資料分布之試驗機差；μ_i 爲品系 i 之平均產量眞值；β_{pi}，β_{fi} 爲品系 i 各在惡劣及有利環境下之穩定性係數眞值；E_j 爲環境 j 之環境指標數值。由上述假設模式，共生產出各200套的「玉米試驗數據」及「水稻試驗數據」，分別以傳統迴歸模式及分段直線迴歸模式進行穩定性分析，其各品系之穩定性分析結果將與迴歸係數眞值互作比較。

爲探討不同迴歸法對品系穩定性係數的估計效率與係數眞值間之關係，本研究由模擬試驗結果，分別計算各迴歸法所估得的各品系穩定性係數估值之相對偏差(relative bias, %)：

bias＝mean of estimate－true value

relative bias（%）＝100×bias／true value

以及相對平方根均方（relative square root of MSE, %）：

root MSE＝〔variance of estimate＋（bias)^2〕$^{1/2}$

relative root MSE（%）＝100×root MSE／true value

結果與討論

一、模擬試驗的比較結果

在玉米假設模式下200套模擬資料均經估算其環境指標估計値後（各環境指標估計值平均與其眞值皆相當接近，標準偏差在50左右），求配傳統迴歸模式及分段直線迴歸模式所得的穩定性分析結果，比較於表 3 。由表中可知，雖然以分段直線迴歸模式所得的穩定性係數之變方均稍大於傳統迴歸法所得者，但顯然前者所得之穩定性係數估計值（b_p 及 b_f）較接近於眞值，而以傳統迴歸法模式所得之穩定性係數（b）則相當於兩個眞值（β_p 及 β_f）的平均。就判斷各品系之適應性所屬類別而言，對於原本在所有環境均表現穩定的品系 1 及 4（第 I 類—生物穩定性）、品系 8（第 V 類—農業穩定性）或不穩定的品系12（第 IX 類），經傳統迴歸法分析結果仍然一致；但倘若基因型在惡劣環境及有利環境的表現有所差異，利用傳統迴歸法分析結果，會發生差異性互相抵銷之現象，當兩者之間的反向差距愈大，則其受到遮蓋效應的影響程度也愈高。從傳統迴歸法分析的結果來看，由於品系 3、5、8、10的穩定性係數估值（b）近於 1，故一般都會認定這些品系應該最符合農業穩定性，但其實不然，品系 3 對惡劣環境最鈍感（$\beta_p=0.2$）而對有利環境則敏感（$\beta_f=1.5$），其次爲品系 5（$\beta_p=0.6$ 及 $\beta_f=1.0$），至於品系10則對惡劣環境最敏感（$\beta_p=1.4$）但對有利環境則鈍感（$\beta_f=0.5$）。玉米模擬資料以分段直線迴歸模式所得的穩定性評估結果，雖然劣境穩定性係數估值（b_p）相對於眞值β_p似乎皆有高估的現象，但仍不至於造成對品系適應性分類的錯誤判斷。由於第 II 及 III 類品系是否爲最理想的基因型，完全決定於平均產量之表現，而由表 3 之各品系平均產量之估值得知，玉米模擬試驗中以品系 9（第 VI 類）之平均產量最高，表示品系 9 在劣境中雖不如第 II 及 III 類品系之表現穩定，但在有利環境下隨環境改善可獲致的產量增進效果卻更大；而品系12（第 IX 類）之平均產量則與品系3, 6（第 III 類）相近，表示品系12雖然對環境變化均相當敏感（$\beta_p=1.4$ 及 $\beta_f=1.5$），但因爲有利環境下的產量增進效果甚大，致足以彌補了其在劣境下的產量損失。由此可知，基因型外表平均值通常與穩定性之間互爲獨立，故必須同時根據平均產量及穩定性係數之表現，才能選拔出農業上眞正理想的基因型。從平均產量估計值相對於眞值的偏差（relative bias, %）來看，當品系對優劣環境表現不一時，對優境敏感之品系的平均產量較易被高估，反之對劣境敏感之品系所估得之平均產量則較爲低估。

表 3. 模擬資料以傳統迴歸及分段直線迴歸所得的穩定性分析之結果比較

Table 3. Comparison between stability analysis using conventional regression and piecewise linear regression obtained by fitting simulated data

Line	Mean yield[w]			Real linear responses[x]			Regression with all environments[y]		Piecewise linear regression[z]			
	μ	m	Rbias %	Category	β_p	β_f	b		b_p		b_f	
							Mean	Var.	Mean	Var.	Mean	Var.
					Maize model							
1	3100	3197	3.1	I	0.2	0.5	0.40	0.008	0.27	0.050	0.52	0.040
2	2700	2955	9.4	II	0.2	1.0	0.68	0.007	0.28	0.054	1.03	0.042
3	2900	3319	14.5	III	0.2	1.5	0.97	0.009	0.32	0.052	1.56	0.040
4	2600	2572	−1.1	I	0.6	0.5	0.61	0.010	0.74	0.052	0.49	0.037
5	2800	2934	4.8	II	0.6	1.0	0.89	0.011	0.76	0.037	1.02	0.037
6	3100	3387	9.3	III	0.6	1.5	1.17	0.009	0.78	0.045	1.53	0.031
7	2800	2641	−5.7	IV	1.0	0.5	0.83	0.010	1.22	0.051	0.48	0.040
8	3200	3197	−0.1	V	1.0	1.0	1.11	0.008	1.25	0.048	0.99	0.038
9	3500	3661	4.6	VI	1.0	1.5	1.39	0.010	1.25	0.058	1.52	0.048
10	2700	2410	−10.7	VII	1.4	0.5	1.03	0.011	1.66	0.051	0.47	0.037
11	2500	2375	−5.0	VIII	1.4	1.0	1.32	0.009	1.70	0.049	0.99	0.039
12	3300	3343	1.3	IX	1.4	1.5	1.58	0.010	1.79	0.048	1.40	0.038
					Rice model							
1	6100	6114	0.2	I	0.3	0.3	0.32	0.005	0.32	0.028	0.31	0.025
2	6400	6695	4.6	II	0.3	1.0	0.70	0.005	0.37	0.030	1.02	0.025
3	6700	7271	8.5	III	0.3	1.7	1.09	0.006	0.41	0.029	1.75	0.027
4	6000	5911	−1.5	I	0.6	0.3	0.46	0.005	0.65	0.033	0.28	0.027
5	6800	6995	2.9	II	0.6	1.0	0.85	0.006	0.69	0.029	1.00	0.026
6	6300	6774	7.5	III	0.6	1.7	1.24	0.005	0.72	0.029	1.73	0.025
7	6000	5807	−3.2	IV	0.9	0.3	0.61	0.006	0.96	0.027	0.27	0.032
8	6300	6379	1.3	V	0.9	1.0	1.00	0.004	1.00	0.028	1.00	0.030
9	6500	6869	5.7	VI	0.9	1.7	1.39	0.005	1.04	0.027	1.73	0.027
10	5700	5397	−5.3	VII	1.2	0.3	0.76	0.005	1.29	0.026	0.26	0.023
11	6300	6280	−0.3	VIII	1.2	1.0	1.14	0.006	1.31	0.031	0.98	0.029
12	5900	6161	4.4	IX	1.2	1.7	1.54	0.006	1.36	0.028	1.71	0.027
13	5400	4994	−7.5	VII	1.5	0.3	0.90	0.006	1.58	0.031	0.25	0.025
14	6200	6078	−2.0	VIII	1.5	1.0	1.29	0.005	1.63	0.026	0.98	0.023
15	6600	6772	2.6	IX	1.5	1.7	1.70	0.005	1.67	0.026	1.73	0.026

[w] μ,m, Rbias% : Real value, mean of estimated value and relative bias (%) of the mean yield for each genotype, respectively ; where relative bias%) =100 $(m-\mu)/\mu$.

[x] Category see Table 1 ; β_p, β_f : True values of regression coefficients in poor environments and favorable environments, respectively.

[y] Mean and variance of the estimated stability indices (b) with all environments from conventional linear regression.

[z] Mean and variance of the estimated stability indices (b_p, b_f) with poor environments and favorable environments, respectively, from piecewise linear regression.

在水稻假設模式下200套模擬資料均經估算其環境指標估計值後（各環境指標估計值平均與其眞值皆相當接近，標準偏差在65左右），求配傳統迴歸模式及分段直線迴歸模式所得的穩定性分析結果比較，亦列於表 3 。由表中可知，以分段直線迴歸法所得的穩定性係數之變方雖稍大於傳統迴歸法所得者，但其所估得之穩定性係數估計值（b_p 及 b_t）較接近於眞值，而以傳統迴歸法所得之穩定性係數（b）則相當於兩個眞值（β_p 及 β_t）的平均；此與玉米假設模式所得結果相同。就判斷各品系之適應性所屬類別而言，對於原本在所有環境均表現穩定的品系 1 及 4（第 I 類一生物穩定性）、品系 8（第 V 類一農業穩定性）或不穩定的品系12、15（第 IX 類），經傳統迴歸法分析結果仍然一致；但對於在惡劣環境及有利環境的表現差異的品系，利用傳統迴歸法分析因爲受到遮蓋效應的影響，有誤判的可能。例如經傳統迴歸法分析結果會認定品系 3、8、13的穩定性係數估值（b）近於 1，應該爲最符合農業穩定性的基因型，但其實上品系 3 對惡劣環境最鈍感（$\beta_p=0.3$）而對有利環境則敏感（$\beta_t=1.7$），品系13則是最不好的基因型，其對惡劣環境最敏感（$\beta_p=1.5$）但對有利環境則鈍感（$\beta_t=0.3$）。水稻模擬資料以分段直線迴歸法所得的穩定性評估結果，雖然所估得之劣境穩定性係數估值（b_p）相對於眞值 β_p 似乎皆有高估的現象，但對於品系適應性分類的判斷，仍與實際反應一致。同時以平均產量及穩定性係數進行選拔，由於品系 3 之產量最高且具高適應性（第 III 類基因型），因此可以認定該品系爲最理想的基因型。以傳統迴歸法所得結果，亦以品系 3 爲最理想基因型（最高產且 b近於 1），但以傳統法並無法清楚瞭解該品系在優劣環境下的表現差異。在水稻模擬試驗結果中亦發現，品系對優劣環境之表現互異，多少會造成平均產量在估計上的偏差，例如品系 3 對優境相當敏感而對劣境則極爲鈍感，因此所受到產量高估的程度也最大。

爲探討不同迴歸模式對品系穩定性係數的估計效率與係數眞值間之關係，本研究綜合上述兩種假設模式模擬試驗結果，進而計算各迴歸法所估得的各品系穩定性係數估值之相對偏差及相對平方根均方，分別列於表 4、5 。從表 4 之相對偏差結果來看，除了眞值 β_p 及 β_t 相近（即品系對所有環境呈線性反應）的情形以外，以傳統迴歸法所得估值的相對偏差均甚大於分段直線迴歸法所得者，表示傳統迴歸法確實不適用於非線性的場合；β_p 及 β_t 的差距愈大，以傳統迴歸法所得估值的相對偏差愈大，如第 III 及 VII 類品系，其估計小於 1 之穩定性係數（即第 III 類之 β_p 及第 VII 類之 β_t）的相對偏差甚至在100%以上；利用傳統迴歸法所得估值的相對偏差，以第 III 類品系所可能發生的相對偏差最大，而該類品系正好是育種上最理想的高適應基因型。利用分段直線迴歸法所得之優境穩定性係數估值的相對偏差均甚小於傳統迴歸法所得者，其絕對值最大不會超過20%；而分段直線迴歸法所得之劣境穩定性係數估值的相對偏差，除了品系對所有環境呈線性反應的場合之外，亦均小於傳統迴歸法所得者，其絕對值最大不會超過50%。由表 5 之相對平方根均方結果而言，則與相對偏差所獲結果一致，只有在眞值 β_p 及 β_t 相近（即品系對所有環境呈線性反應）的場合，傳統迴歸法才有較好的估計效率，當 β_p 及 β_t 的差距愈大，則估計效率愈差，其中對於最理想品系（第 III 類）及最差品系（第 VII 類）的選拔最容易產生誤判。顯然，只有在品系呈非線性反應的場合，利用分段直線迴歸法的估計效率，才有可能高於傳統迴歸法，倘絕大多數品系對所有環境呈線性反應，則分段直線迴歸法與傳統迴歸法的表現不相上下（表 5）。

表 4. 以傳統迴歸及分段直線迴歸評估基因型穩定性係數的相對偏差

Table 4. The relative bias in estimating the stability indices of genotypes using conventional regression and piecewise linear regression

Category[w]	Real linear responses[x]		Relative bias (%)			
			Poor environments		Favorable environments	
	β_p	β_f	b^y	b_p^z	b	b_f^z
I	0.2	0.5	100.11	30.81	−19.96	4.38
	0.3	0.3	7.01	11.26	7.01	2.72
	0.6	0.3	−23.81	7.94	52.38	−8.15
	0.6	0.5	3.01	22.09	23.61	2.30
II	0.2	1.0	236.83	47.58	−32.63	1.25
	0.3	1.0	135.43	24.46	−29.37	2.06
	0.6	1.0	45.05	19.98	−12.97	1.06
III	0.2	1.5	379.20	50.74	−36.11	3.38
	0.3	1.7	264.34	36.29	−35.71	2.44
	0.6	1.5	96.50	30.53	−21.40	2.36
	0.6	1.7	106.60	20.88	−27.08	1.59
IV	0.9	0.3	−32.00	7.10	103.99	−7.45
	1.0	0.5	−17.38	20.15	65.24	−2.31
V	0.9	1.0	12.05	12.91	0.85	0.14
	1.0	1.0	10.60	20.77	10.60	1.43
VI	0.9	1.7	54.66	16.54	−18.12	1.02
	1.0	1.5	38.70	26.26	−7.54	−0.01
VII	1.2	0.3	−37.17	5.65	151.33	−11.05
	1.4	0.5	−25.52	20.19	108.55	−6.44
	1.5	0.3	−39.85	6.67	200.76	−19.57
VIII	1.2	1.0	−4.24	10.06	14.92	−1.32
	1.4	1.0	−5.92	22.48	31.72	−3.91
	1.5	1.0	−14.03	7.18	28.96	−1.15
IX	1.2	1.7	28.01	12.19	−9.64	1.00
	1.4	1.5	13.99	27.35	6.39	−4.70
	1.5	1.7	12.97	11.59	−0.32	0.83

[w] See Table 1.

[x,y,z] See Table 3.

表 5. 以傳統迴歸及分段直線迴歸評估基因型品系穩定性係數的相對平方根均方

Table 5. The relative root MSE in estimating the stability indices of genotypes using conventional regressing and piecewise linear regression

Category[w]	Real responses[x]		Relative root MSE (%)			
			Poor environments		Favorable environments	
	β_p	β_f	b^y	b_p	b	b_f^z
I	0.2	0.5	112.46	114.49	28.61	42.62
	0.3	0.3	28.69	61.78	28.69	59.70
	0.6	0.3	27.16	31.90	58.53	61.03
	0.6	0.5	17.52	46.01	31.41	44.08
II	0.2	1.0	241.87	125.52	34.08	20.92
	0.3	1.0	137.62	61.30	30.27	16.56
	0.6	1.0	47.37	39.43	15.84	18.68
III	0.2	1.5	382.50	128.62	36.72	14.70
	0.3	1.7	265.62	69.46	36.00	10.11
	0.6	1.5	98.22	49.07	22.62	14.11
	0.6	1.7	107.30	35.74	27.43	10.11
IV	0.9	0.3	32.99	20.56	106.74	54.01
	1.0	0.5	20.07	30.95	68.26	43.22
V	0.9	1.0	14.39	23.26	7.13	17.59
	1.0	1.0	14.62	32.20	14.62	22.72
VI	0.9	1.7	55.27	25.49	18.63	9.31
	1.0	1.5	40.11	35.51	10.30	14.09
VII	1.2	0.3	37.69	16.10	153.41	60.61
	1.4	0.5	26.59	26.74	110.54	46.05
	1.5	0.3	40.20	13.94	202.51	56.01
VIII	1.2	1.0	7.00	18.22	16.35	16.20
	1.4	1.0	9.53	28.48	33.40	23.99
	1.5	1.0	14.91	13.10	29.94	16.66
IX	1.2	1.7	28.76	18.73	10.68	9.27
	1.4	1.5	15.48	32.04	8.90	14.63
	1.5	1.7	13.93	16.48	4.48	9.37

[w,x,y,z] See as Table 4.

三、實例應用分析及 SAS 程式說明

本研究用以進行實例應用分析的資料來源爲1986年至1987年之本省 5 個飼料玉米綜合品系在 3 個地區之春作籽實產量區域試驗資料[(9)]，該試驗方法採逢機完全區集設計，重複 4 次，依本省雜糧作物育種實施程序，一般以年度及地區之組合爲環境變異，故合計 5 個品系 6 個環境 4 個重複進行穩定性分析。

分段直線迴歸模式之穩定性分析的 SAS 程式，列於附錄 1。程式中被兩個/*符號所包含於內的文字，爲程式說明，純粹供輔助讀者瞭解各程序步驟之用，SAS 並不會執行，因此在實際輸入時可以除去無妨。程式內容之主要步驟依次爲：資料讀入、綜合變方分析之執行、品系平均值之計算、總平均值之計算、每一品系在各環境下所有重複之平均值之計算、環境指標之估算、環境指標之劃分、指示變數之定義、分段直線迴歸分析之進行、優劣境穩定性係數之估算、以及統計值摘要表之建立。該程式在應用不同試驗資料時，只須更改資料讀入部份中第 2 行敍述的 INFILE 之輸入資料檔名，並決定第 3～5 行敍述的品系數（K）、環境數（N）及重複數（L）即可，如本文實例試驗資料檔爲 plrstab.dat，K=5，N=6，L=4。資料檔係以先讀重複後，再讀環境，最後讀品系的輸入格式建立，倘欲將資料輸入格式改以先讀重複後再讀品系後讀環境的方式，則必須將程式中的資料輸入迴圈（DO-Loops）更改爲 DO ENV=1 TO N；DO LINE=1 TO K；DO REP=1 TOL。除此，該程式之所有步驟中的敍述或數值，均毋須再作任何更動。當試驗沒有重複時（L=1），該程式亦能順利執行，惟所輸出的綜合變方分析中的機差項將無法估計，但無礙於之後的穩定性分析。程式之輸出結果內容包含有綜合變方分析、各品系在不同環境下之產量平均值、環境指標估值、各品系對環境指標之分段直線迴歸分析、穩定性分析結果之摘要表（列舉各品系之平均產量、b_p 及 b_f 估計值）。

玉米區域試驗資料分別利用傳統直線迴歸模式及分段直線迴歸模式進行穩定性分析的結果，列於表 6。由表 6 發現，所有品系配合傳統法及分段直線迴歸法的 F 檢定值均顯著或極顯著，且其模式的滿意程度（R^2）均甚高，即迴歸關係皆顯著存在，同時兩種方法所有品系之迴歸離差均方皆互爲均質，表示該試驗資料能滿足直線迴歸法應用於穩定性分析的兩個重要前提(5)，故兩種迴歸法所估得之迴歸係數均足以解釋基因型對環境的適應能力。分段直線迴歸法較傳統迴歸法所得之 R^2 爲高，且其中以品系 PCC 增加幅度最大（由0.86提高爲0.91），表示該品系對環境的反應較偏向於非線性關係。根據傳統迴歸法所估得的穩定性係數結果，四個參試品系之 b 值均相對地較對照品系爲低，即表現可更爲穩定，其中 PC8 及 PC7 之 b 值最接近於 1，故一般認爲是最符合農業觀念的穩定品種，而 PCC 及 PUC 之 b 值爲0.81，小於 1，對環境的變異亦相當穩定，尤其在惡劣環境下可較 PC8 及 PC7 穩定，但因環境改善所能提高的產量卻不如 PC8 或 PC7 爲高。倘若利用分段直線迴歸法進行穩定性分析，則可更清楚觀察出各品系對優劣環境下的表現差異。由於農業上多同時以穩定性測量值與產量平均值來作爲選拔高產且穩定品種之依據，故將所估得的劣境穩定性係數(b_p)、優境穩定性係數（b_f）及平均產量值三者之間的關係繪成圖示，則有利於基因型適應性的分類及篩選，在參試品系數甚多的場合尤其有用。如圖 1 所示，根據 b_p 及 b_f 兩座標軸可將所有品系區分爲如表 1 所述之九種類別的適應性。各品系的產量平均值，則可由各品系垂直線的高度來進行比較。由本研究玉米區域試驗結果（圖 1），PUC 與對照品系的適應性屬於同一群（第Ⅲ類），雖爲具高適應性的品系，其對劣境鈍感而對優境則敏感，但 PUC 的產量平均值卻不如對照品系，也未高於產量總平均值（2482 kg ha^{-1}）；PCC 的適應性最差，其對劣境敏感而對優境鈍感（第Ⅶ類），且產量平均值也最低，爲最不好的基因型；PC8 及PC7 的適應性相近（第Ⅷ類），雖然對環境改善能獲得產量增進（b_f 近於1），但對惡劣環境也相當敏感（$b_p>1$），不過因爲 PC8 的產量優於對照品系，表示在本試驗環境範圍內 PC8 因處於有利環境所獲得的產量增進，足以彌補其在劣境下的產量損失；但仍然必須留意該品系往後可能受到不利環境所引起的低產風險。

表 6. 五個玉米品系在六種環境之區域試驗資料以傳統迴歸及分段直線迴歸法進行穩定性分析之結果比較

Table 6. Comparison between stability analysis using conventional regression and piecewise linear regression in regional trial data of five maize composites under six environments

Variety	Mean yield	Regression with all environments				Piecewise linear regression				
	(kg ha^{-1})	F-test	R^2	b	S^2_d	F-test	R^2	b_p	b_f	S^2_d
PC8	2743	**	0.97	1.10	41879	**	0.98	1.41	0.81	432023
PC7	2589	**	0.93	1.11	96429	*	0.93	1.32	0.93	123393
PCC	2014	**	0.86	0.81	107299	*	0.91	1.44	0.26	93438
PUC	2369	**	0.92	0.81	92001	*	0.93	0.19	1.36	73961
CK[z]	2696	**	0.91	1.17	136976	*	0.93	0.65	1.64	147577
Grand mean	2482									
Bartlett's x^2 test					ns					ns

[z] Tainan sel No. 10.
F-test : Significant test of F values for regression model.
R^2 : Coefficient of determination.
S^2_d : Mean squares of deviation from regression.
b : Estimated stability indices with all environments from conventional linear regression.
b_p,b_f : Estimated stability indices with poor environments and favorable environments respectively, from piecewise linear regression.
*,** : Significant at 5% and 1% level, respectively.
ns : Non-significant.

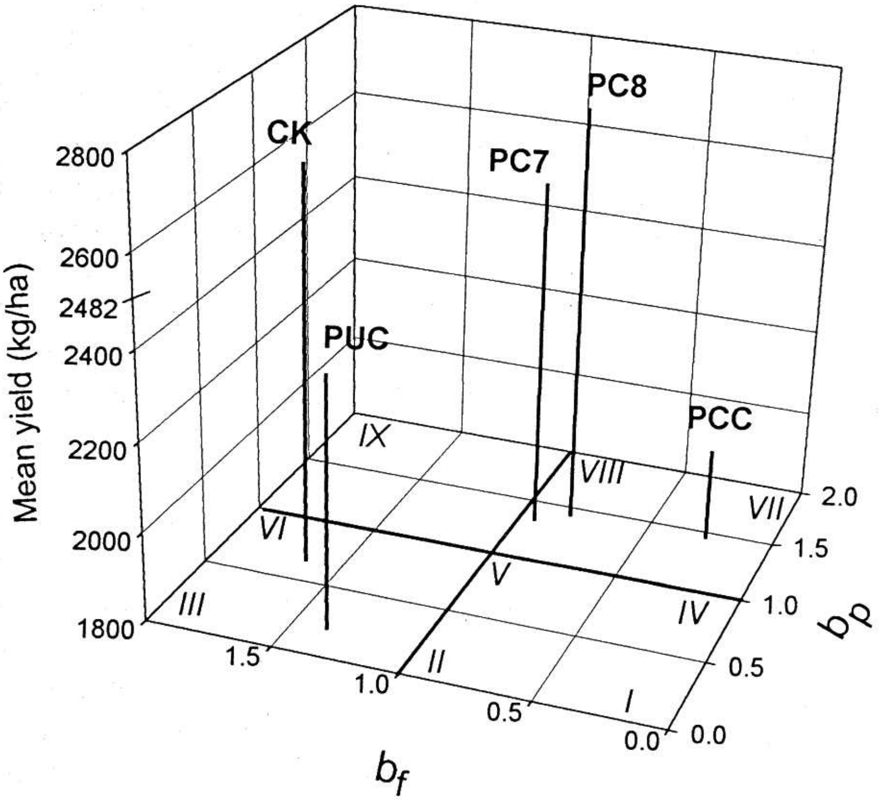

圖 1. 玉米各品系之劣境穩定性指標 (b_p) 、優境穩定性指標 (b_f) 及平均產量間的關係

Fig. 1. The relationship among stability index with poor environments (b_p), stability index with favorable environments (b_f) and mean yield for each maize genotype.

討　論

Verma and Chahal 之分段直線迴歸模式，係針對傳統直線迴歸穩定性分析所提出的修正迴歸法，對評估廣大環境範圍之適應性，在觀念上頗爲實用，尤其在基因型對環境呈非線性反應的場合[13,14]。從本研究模擬試驗之結果，也證實了傳統迴歸模式在評估作物對環境之非線性反應的不適用性，而利用分段直線迴歸模式可以有效地檢測出品系對優劣環境的反應差異；但倘若品系對所有環境呈線性反應，則利用分段直線迴歸法的估計效率未必就高於傳統迴歸法。如果非線性反應存在，分段直線迴歸法是否就能完全取代複雜的曲線模式，仍言之過早，有待再進一步深入研究探討。其癥結問題所在，是應該找出從直線變成曲線的結點是發生在哪一個環境範圍內，而各品系對環境反應的形式應該不同，即各品系的結點未必就會發生在同一點上，如 Verma and Chahal 之將結點設定在 0，係假設所有品系均在環境指標數值等於 0 時從直線變成曲線，這樣的假設在實際狀況下並不常見，以致 Verma and Chahal 之分段直線迴歸模式的應用性迄今仍相當有限。再者，以栽培者的標準來說，田間研究中所謂的劣境，並不見得就是不利於作物生長的環境；在分析中或許可以武斷地將環境半數歸類爲劣境，由於實在很難確定參試環境就能適當地代表整個環境族群，如果品系生長的環境包含了所有的環境地區，則將會因爲環境指標估計值與研究中的地區平均值不同，穩定性評估結果便有很大的差距，因此在使用分段直線迴歸模式於穩定性分析時，除了利用環境指標之正負值將環境以武斷方式等分之外，或許也可憑著經驗或知識將環境分爲劣、優兩類。Lin *et al.*[5] 也認爲迴歸分析中之環境指標若能以實際的外在氣候因子來取代，則有利於發展爲預測模式。惟利用外在變數來描述環境，固然有很多優點，但在很多場合，環境的生產水準不可能單純地只由一些有限的環境因子所決定[1]，利用外在變數以預測交感（特別是產量之交感）至今仍證實相當棘手[6]。由於基因型外表型平均值與穩定性之間通常互爲獨立，故必須同時考慮兩者才能選出高產且高適應性的優良品種。這樣的基因型是否存在？是尚無法驗證可得到答案的問題。因爲基因型外表型平均值與穩定性爲數量性狀，控制這些性狀的基因或基因集幾乎都在不同的染色體上[2]，故欲育成高產且高適應性之品種，實際上相當困難[6]。

附錄 1. 分段直線迴歸穩定性分析之 SAS 程式

Appendix 1. The SAS program for piecewise linear regression for stability analysis

```
DATA PLRSTAB;
INFILE "plrstab.dat" ;
K=5;
N=6;
L=4;
DO LINE=1 TO K;          /*A two-way table with K lines × N environments × L replicates*/
  DO ENV=1 TO N;
     DO REP=1 TO L;
       INPUT YIELD @@;
       OUTPUT;
     END;
   END;
END;
TITLE ' -- Combined Analysis of Variance -- ';
```

(續附錄 1.)

```
PROC ANOVA OUTSTAT=STAT ;                                  /*Performing Combined ANOVA*/
  CLASS REP ENV LINE ;
  MODEL YIELD=REP(ENV) ENV LINE LINE*ENV ;
  TEST H=ENV E=REP(ENV) ;
PROC MEANS MEAN NOPRINT DATA=PLRSTAB ;                     /*Calculation of line mean*/
  VAR YIELD ;
  BY LINE ;
  OUTPUT OUT=L MEAN=LM ;
PROC MEANS MEAN NOPRINT DATA=PLRSTAB ;                     /*Calculation of grand mean*/
  VAR YIELD ;
  OUTPUT OUT=L MEAN=GM ;
PROC MEANS MEAN NOPRINT DATA=PLRSTAB ;                     /*Calculation of replicate mean*/
  VAR YIELD ;
  BY LINE ENV ;
  OUTPUT OUT=LE MEAN=LEM ;
PROC PRINT NOOBS ;
  TITLE '-- Line means in each environment (LEM) --' ;
  VAR LINE ENV LEM ;
PROC SORT DATA=LE ;
  BY ENV ;
PROC MEANS MEAN NOPRINT ;                                  /*Estimation of environmental index*/
  BY ENV ;
  VAR LEM ;
  OUTPUT OUT=E MEAN=EM ;
DATA ENVIND ;                                              /*Partition of environmental index*/
  IF_N_=1 THEN SET M ;
  SET E ;
  EI=EM-GM ;
  IF EI<=0 THEN INDEX=0 ;                                  /*Defining the indicator variable*/
  IF EI>0 THEN INDEX=1 ;
  EI2=EI*INDEX ;
TITLE ' -- Environmental index --' ;
PROC PRINT NOOBS LABEL ;
  LABEL EI='ENV. INDEX' ;
  VAR ENV EI ;
DATA COMBINE ;
  MERGE LE ENVIND ;
  BY ENV ;
  DROP _TYPE_ _FREQ_ ;
PROC SORT OUT=COMBINE ;
  BY LINE ;
TITLE ' -- Piecewise regression analysis of each line --' ;
PROC REG OUTEST=EST ;                      /*Piecewise regression of yield on environmental index*/
```

(續附錄 1.)

```
  BY LINE ;
  MODEL LEM=EI EI2 ;
DATA ESTB ;
  SET EST ;
  KEEP LINE EI EI2 ;
DATA OUTLIST ;
  MERGE ESTB L ;
  BY LINE ;
  DROP _TYPE_ _FREQ_ ;
  EP=EI ;                                   /*Estimation of two stability indices*/
  EF=EI+EI2 ;
TITLE ' -- Summary of mean yield and stability index --' ;
PROC PRINT NOOBS LABEL ;                    /*List of mean yield and stability indices*/
  LABEL LINE='LINE'
        LM='YIELD MEAN'
        EP='b in poor environments'
        EF='b in favorable environments' ;
        VAR LINE LM EP EF ;
RUN ;
```

引用文獻

1. Becker, H. C. and J. Lēon. 1988. Stability analysis in plant breeding. Plant Breed. 101 : 1—23.
2. Berke, T. G., P. S. Baenziger and R. Morris. 1992. Chromosomal location of wheat quantitative trait loci affecting stability of six traits, using reciprocal chromosomal substitutions. Crop Sci. 32 : 628—633.
3. Eberhart, S. A. and W. A. Russell. 1966. Stability parameters for comparing varieties. Crop Sci. 6 : 36—40.
4. Finlay, K. W. and G. N. Wilkinson. 1963. The analysis of adaptation in a plant breeding programme. Austral. J. Agric. Res. 14 : 742—754.
5. Lin, C. S., M. R. Binns and L. P. Lefkovitch. 1986. Stability analysis : Where do we stand? Crop Sci. 26 : 894—900.
6. Lin, C. S. and M. S. Binns. 1994. Concepts and methods for analyzing regional trial data for cultivar and location selection. Plant Breed. Reviews 12 : 271—297.
7. Lu, H. Y. 1988. Stability analysis. Sci. Agric. 36 : 333—339. (in Chinese)
8. Lu, H. Y., C. T. Lu and C. C. Chen. 1992. Effects of yield components and agronomic traits on yield stability of Japonica rice. J. Agric. Res. China 41 : 317— 330. (in Chinese with English summary)
9. Lu, H. Y. and C. T. Tien. 1993. Studies on nonparametric method of phenotypic stability. II. Selection for stability of agronomic concept. J. Agric. Assoc. China 164 : 1—16. (in Chinese with English summary)
10. Lu, H. Y. 1994. A SAS program for regression analysis of stability. J. Agric. Res. China 43 : 283—292. (in Chinese with English summary)
11. Neter, J., M. H. Kutner, C. J. Nachtsheim and W. Wasserman. 1996. Applied Linear Regression Models. Richard D. IRWIN, 3rd, London. pp. 474—477.

12. Perkins, J. M. and J. L. Jinks. 1968. Environmental and genotype-environmental components of variability : III. Multiple lines and crosses. Heredity 23 : 339—356.

13. Pritts, M. and L. Luby. 1990. Stability indices for horticultural crops. Hort Science 25 : 740—745.

14. Verma, M. M. and G. S. Chahal. 1978. Limitations of conventional regression analysis : A proposed modification. Theor. Appl. Genet. 53 : 89—91.

15. Westcott, B. 1986. Some methods of analysing genotype-environment interaction. Heredity 56 : 243—253.

16. Wright, A. J. 1976. Bias in the estimation of regression coefficients in the analysis of genotype-environment interaction. Heredity 37 : 299—303.

17. SAS Institute Inc. 1990. SAS/STAT User's Guide. Version 6, 4th ed., Cary, NC.

Comparative Study of a Piecewise Linear Regression Model with the Conventional Regression Model for Assessing Nonlinear Environmental Responses of Crops[1]

Hsiu-Ying Lu[2]

Summary

A high adaptable crop variety would be the one with relatively low sensitivity in the poor environments and high sensitivity in the favorable environments. That is, an ideal genotype should have nonlinear responses to environment. The conventional linear regression for stability analysis cannot detect such varieties, and an unstable variety might be wrongly regarded as stable, because the computation of the linear regression over all the environments has the masking effect on the detection of ideal genotypes, if they exist in the population. The purpose of this study was to compare the validity of the piecewise linear regression model proposed by Verma and Chahal (1978) and the conventional regression model in stability analysis. Based on hypothetical models of maize and rice, two hundred sets of simulated data generated from each hypothetical model were analyzed by each of the conventional regression and piecewise linear regression. The result showed the piecewise linear regression is useful to measure the separate responses of varieties to the defferent ranges of environments, and reduce the risk of wrongly identification. A SAS program for the piecewise linear regression analysis of stability was presented. This program runs on personal computers. The procedure was illustrated through an example of maize regional trial data in Taiwan. It is useful to agricultural researchers in application of piecewise linear regression for stability analysis. The relationship among stability in poor environments, stability in favorable environments and mean performance for each genotype was shown by a three-dimensional graph. This graph provides a direct and easy method of screening genotypes with the performance above the mean and high adaptability.

Key words : Stability, Piecewise linear regression, Nonlinear responses, Simulation experiment, Hypothetical model, SAS program.

1. Contribution No. 1927 from Taiwan Agriculture Research Institute.
2. Statistician, Department of Agronomy, TARI, Wufeng, Taichung, Taiwan, ROC.